Manuscript: Annual flood sensitivities to El Niño Southern Oscillation at the global scale
Authors: P. J. Ward, S. Eisner, M. Flörke, M. D. Dettinger, and M. Kummu
Response to Reviewer #2

General response points

We thank the reviewer for the very positive comments and are very pleased that he/she values the scientific relevance of our research. The reviewer provides several very useful comments/suggestions for minor revisions. We will address these in the revised manuscript, as per our responses to each comment below.

Specific points

1. Can the authors provide some justification for selecting the basin outlet cell to represent flood timing / quantity for the full basin (or sub-basin.) Physically it is logical, but how well does the WaterGap model perform on a gridded basis? Are the sub-basins relatively homogeneous from a gridded perspective?

Indeed, we had several discussions on whether to present the results of the analyses per grid-cell or per sub-basin. The decision to report the results per basin is mainly based on the fact that we can expect global model results to be more erroneous in upstream areas containing (for example) just one grid cell, rather than areas towards the outlet where the discharge has been routed over a larger area. This is not only to do with the performance of hydrological models (or WaterGAP in particular) but also due to the input climate data. It is well known that climatological input data from reanalysis products (and other products) are subject to very large errors in individual cells. However, it is useful to point out that we did in fact also carry out the analyses at the grid-cell level. Hence, we are here able to show (in a visual sense) that sub-basins results are (relatively) homogenous compared to the gridded results: see the figure below where we show the sensitivity of Q_{max} *to SOI for at the (a) gridded scale, and (b) basin scale. We will add the following line to the manuscript justifying carrying out the analyses at the basin scale. "We present the results at this scale because errors in upstream areas containing just one or a small number of grid cells may be large, particularly due to the use of coarse (0.5° x 0.5°) forcing climate data."*

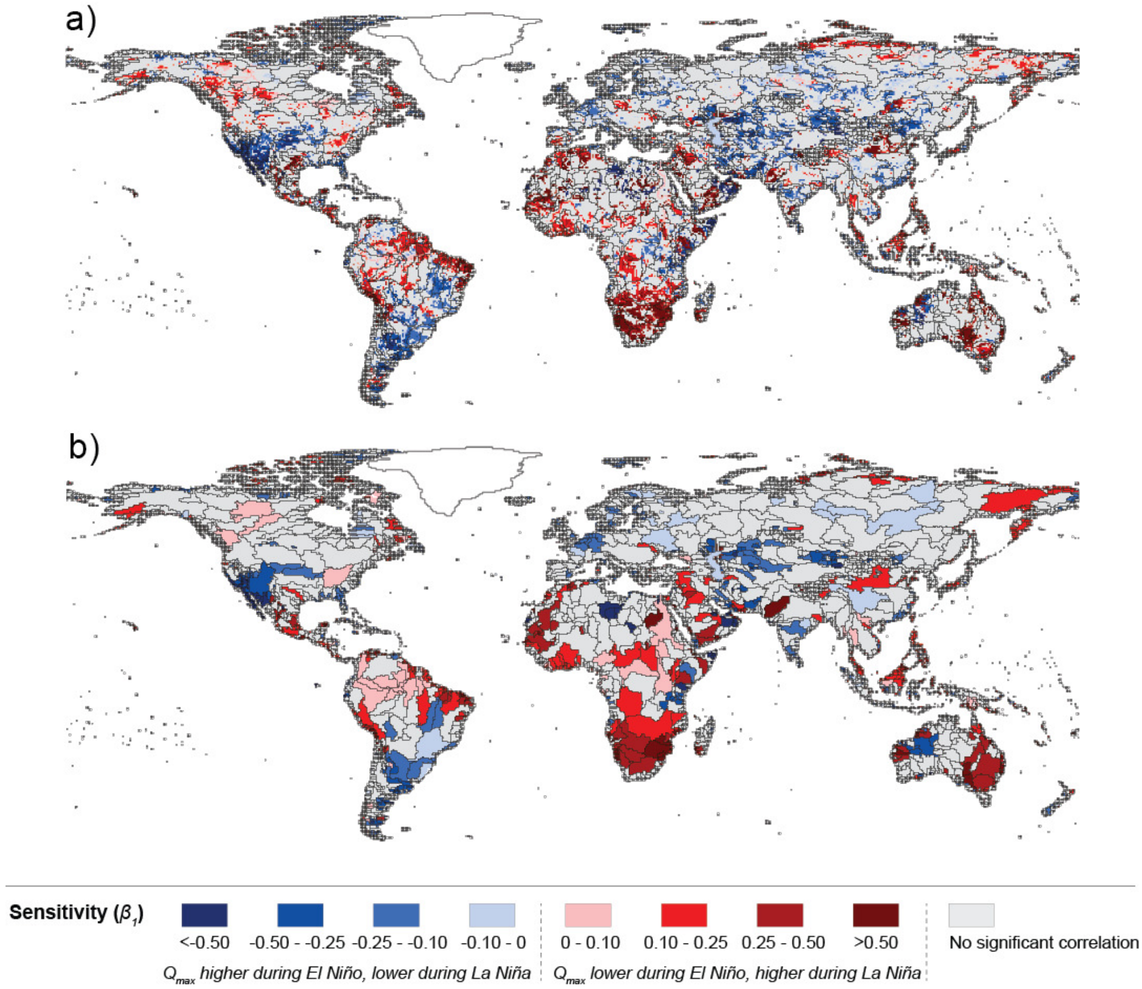

Figure: Sensitivity (β1) of ln Q_{max} to variations in SOI, for basins with significant correlation at a 10% confidence interval (Pearson's r, t-statistic, α = 0.10) (basins where the correlation is not significant are shown in grey). In the upper plate (a), the sensitivity is shown per grid-cell, and in the lower plate (b) the results are shown per basin, based on the value at the outlet cell. Blue indicates negative correlation (higher annual floods in El Niño years/lower annual floods in La Niña years); and red indicates positive correlation (lower annual floods in Niño years /higher annual floods in La Niña years).

2. When the first batch of results are presented (e.g. correlations between SOI and Qmax), it is not clear which season (or 3 month period) for SOI is selected, first referenced on P 10237, Line 5. Is it DJF like the ensuing results section?

For these analyses we actually show the sensitivity (of lnQ_{max}) to 3-monthly averages of the SOI for the month OND, NDJ, DJF, and JFM), where for each basin the sensitivity shown is for the 3-monthly period of the SOI that is most highly correlated to lnQ_{max}. This is stated on P10240, from L8. Hence, the 3-month period is not the same for all basins. We chose to do this because it gives more information on the correlations than simply one fixed 3-month term, since for different basins a different 3-month term may be of interest. However, it may be useful for the reader to also see the results based on the individual 3 month series, to gain an understanding of which 3-month period shows the strongest correlation in each basin. Hence, we will add the following figure (shown below) to the revised manuscript in the supplementary information section, along with a short description of the key differences. In short, the sensitivities are generally similar *between the periods OND, NDJ, and DJF, but the strength of the correlations breaks down in the majority of regions by JFM (which is why periods after JFM were not used).*

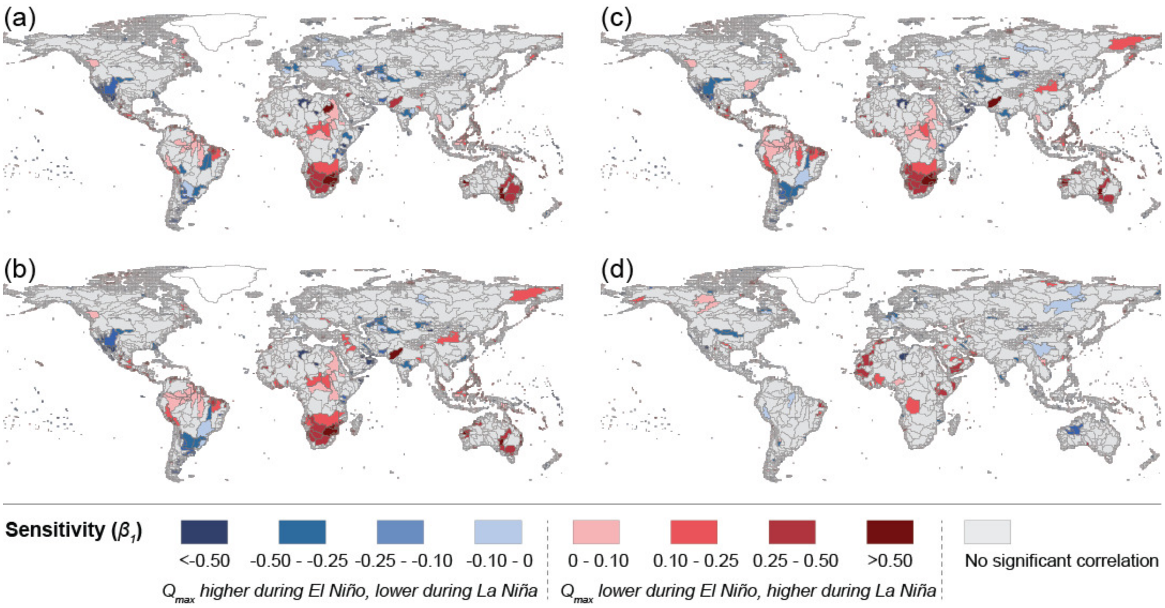

Figure: Sensitivity (β1) of ln Qmax to variations in SOI for basins with significant correlation (Pearson's r, t-statistic, α = 0.10) (basins where the correlation is not significant are shown in grey). The sensitivity is shown for hydrological year Q_{max} to 3-month mean SOI for: (a) October-November-December (OND); (b) November-December-January (NDJ); (c) December-January-February (DJF); and (d) January-February-March (JFM). Blue indicates negative correlation (higher annual floods in El Niño years/lower annual floods in La Niña years); and red indicates positive correlation (lower annual floods in Niño years /higher annual floods in La Niña years).

3. Little is mentioned about evaluating the skill in various 3 month periods. Was this performed, or was a single 3 month period selected for all evaluations?

Indeed, please see response to comment #2 (above).

4. On P 10242, last paragraph, the authors mention high sensitivity in Qmax to ENSO for arid regions. Could this also be a function of the low absolute values of discharge? For example, a doubling in discharge may not necessarily be surprising if mean discharge is relatively low.

It may be the case that the anomalies in peak discharge between the ENSO phases (such as those shown in Figure 8) are higher in arid regions as a result of this. However, for the sensitivities for these analyses we did not use discharge anomalies, but rather absolute values of (log) Q_{max} per year.

With regards the anomalies, there could be a whole range of explanations, including that one suggested. In some regions, another possible explanation could be that the large anomalies could be related to the fact that these biggest (annual) floods reflect a regime where most of the precipitation (and flows) come in just a very few isolated storms; thus all ENSO has to do is to modify the number of storms or the magnitude of a few storms, and it can have a large influence on any given year's Q_{max}. In other settings where the number of storms is larger (more wet days overall), ENSO may need to have a much larger, more persistent, influence to be reflected in Q_{max}.

Another possibility is related to the characteristics of arid areas themselves: arid regions are particular challenging for hydrological simulations, with particular characteristics such as high rainfall variability, extensive surface runoff, reduction of infiltration capacity (crusted soils), lack of vegetation, and so on.

We therefore feel that further research would be required to give a sound answer to this question, and therefore add the amended passage: "However, less research has assessed the influence of ENSO on the hydroclimatology in arid regions. Whilst the paucity of observed discharge data in many of these regions limits the validation of our model results there, the strength of the signal provides motivations for enhancing research activities in those regions, in order to examine whether this is related to physical processes (and if so which), and/or whether this is related to the high coefficient of variability in peak flows".

5. Are 21 windows enough data points to confidently justify (or even evaluate) the strength of the ENSO relationship with Qmax? And enough to claim non-stationarity? There may also be other modulating factors happening concurrently that are less tied to ENSO (e.g. local features.) This may be worth mentioning at the least.

We made a trade-off between the number of years per window (21) and the number of windows (21), in order to try to preserve power in both the correlation test (Pearson's r) and the test of trend (Mann-Kendall). We do believe that the 21 values therefore available for the Mann-Kendall test are sufficient to detect a trend, and therefore non-stationarity, as can also be seen by the fact that statistically significant trends are indeed detected for some basins. However, we have now also tested how stable these trends are if we use a shorter window. We repeated the analyses using a moving window of 15 years, which therefore gave us 27 "windows" from which to detect the trend. Based on this, we found only a small number of changes, namely the Yellow River, Murray, and Ohio displayed no significant trend (instead of strengthening), and the Tocantins displayed no significant trend (instead of weakening). These changes do not affect the overall storyline and conclusions of the paper, and it should be noted that the 15-year time period used to calculate the individuals values of Pearson's r is short. Hence, we choose to keep the 21-year moving window. We will add the following statement to explain why this was chosen: "A 21-year moving window was used as a trade-off to maximise both the number of years per window (21) and the number of windows (21).". Moreover, we will also add the following text "We also repeated the analyses using a 15-year moving window (which yields 27 windows for the trend detection). The results of the latter analyses were similar to those using a 21-year moving window, with the following differences: the Yellow, Murray, and Ohio rivers displayed no significant trend (instead of strengthening), and the Tocantins displayed no significant trend (instead of weakening)." We also considered testing the results using a 30-year moving window. However, this yields just 12 windows, which we consider to be too short for a meaningful assessment of the trend.

With regards the latter part of the comment, i.e. that any apparent changes in correlation strength i.e. non-stationarities) may not necessarily be tied to changes in ENSO, but rather other factors, we agree fully with reviewer. Our data cannot provide any conclusive evidence on whether the non-stationarities are caused by ENSO modulations per se. In fact, we do mention this in the manuscript P10245, L22-29: "On the whole, of the 35 basins highlighted in Fig. 7, *correlations strengthened in 14 basins, weakened in 13, and exhibited no trend in 8. Thus, globally, there has been essentially no overall bias among the changing teleconnections in one direction or the other. This even global mix of strengthening vs. weakening teleconnections may suggest that the changes shown in Figs. 6 and 7 reflect changes in teleconnection strengths, rather than changes in the strength of the driving ENSO variations. The latter might more likely have yielded more universally consistent changes in flood correlations.". What may be causing these "changes in teleconnection strengths, rather than changes in the strength of the driving ENSO variations" is a matter for further research, but indeed could be related to more local/regional features.*